

Expéditeur : le BUREAU INTERNATIONAL

**PCT**

NOTIFICATION RELATIVE  
À LA TRANSMISSION DE LA COPIE DE  
LA DEMANDE INTERNATIONALE TELLE  
QUE PUBLIÉE OU PUBLIÉE À NOUVEAU

Destinataire :

LOUISET, Raphaël  
Cabinet Plasseraud  
65/67, rue de la Victoire  
F-75440 Paris Cedex 09  
FRANCE

**REÇU LE**

**18 JAN. 2005**

**Gbt Plasseraud**

Date d'expédition (jour/mois/année)  
13 janvier 2005 (13.01.2005)

Référence du dossier du déposant ou du mandataire  
BCT040093/AH

**IMPORTANT NOTICE**

Demande internationale n°  
PCT/FR2004/001493

Date du dépôt international (jour/mois/année)  
16 juin 2004 (16.06.2004)

Date de priorité (jour/mois/année)  
25 juin 2003 (25.06.2003)

Déposant

FRANCE TELECOM etc

Le Bureau international transmet ci-joint les documents suivants :

- ☒ copie de la demande internationale telle que publiée par le Bureau international le 13 janvier 2005 (13.01.2005) sous le n° WO 2005/004002
- ☐ copie de la demande internationale telle que publiée à nouveau par le Bureau international le sous le n° WO
- Pour toute explication concernant la nouvelle publication de la demande internationale, se reporter aux codes INID (15), (48) ou (88) (selon le cas) indiqués sur la page de couverture du document joint.

BEST AVAILABLE COPY

Bureau international de l'OMPI  
34, chemin des Colombettes  
1211 Genève 20, Suisse

Fonctionnaire autorisé

Athina Nickitas-Etienne

n° de télécopieur+41 22 740 14 35

n° de télécopieur+41 22 338 89 95

Procédé de traitement d'une séquence sonore, telle qu'un  
morceau musical

La présente invention concerne le traitement d'une  
5 séquence sonore, telle qu'un morceau de musique ou, de  
façon plus générale, une séquence sonore comportant la  
répétition d'une sous-séquence.

Les distributeurs de productions musicales, par exemple  
10 enregistrées sur CD, cassette ou autre support, mettent à  
la disposition des clients potentiels des kiosques où les  
clients peuvent écouter une musique de leur choix, ou  
encore des musiques promues en raison de leur nouveauté.  
Lorsqu'un client reconnaît un couplet ou un refrain du  
15 morceau musical qu'il écoute, il peut décider d'acheter la  
production musicale correspondante.

De façon plus générale, un auditeur d'attention moyenne  
concentre davantage son attention sur un enchaînement de  
20 couplet et refrain, que sur l'introduction du morceau,  
notamment. On comprendra ainsi qu'un résumé sonore  
comportant au moins un couplet et un refrain suffirait  
pour être diffusé dans des kiosques du type précité,  
plutôt que prévoir la production musicale complète à  
25 diffuser.

Dans une autre application telle que la transmission de  
données sonores par téléphonie mobile, on comprendra que  
le téléchargement du morceau de musique complet sur un  
30 terminal mobile, à partir d'un serveur distant, est

beaucoup plus long et, de là, plus onéreux que le téléchargement d'un résumé sonore du type précité.

De même, dans un contexte de commerce électronique, des  
5 résumés sonores peuvent être téléchargés sur un poste  
communiquant avec un serveur distant, via un réseau étendu  
du type INTERNET. L'utilisateur du poste informatique peut  
ainsi passer commande d'une production musicale dont il  
apprécie le résumé sonore.

10

Toutefois, détecter à l'oreille un couplet et un refrain  
et créer ainsi un résumé sonore pour toutes les  
productions musicales distribuées serait une tâche d'une  
lourdeur prohibitive.

15

La présente invention vient améliorer la situation.

L'un des buts de la présente invention est de proposer une  
détection automatisée d'une sous-séquence répétée dans une  
20 séquence sonore.

25

Un autre but de la présente invention est de proposer une  
création automatisée de résumés sonores du type décrit ci-  
avant.

30

A cet effet, la présente invention porte d'abord sur un  
procédé de traitement d'une séquence sonore, dans lequel :  
a) on applique une transformée spectrale à ladite séquence  
pour obtenir des coefficients spectraux variant en  
fonction du temps dans ladite séquence.

Le procédé au sens de l'invention comporte en outre les étapes suivantes :

- b) on détermine, par analyse statistique desdits coefficients spectraux, au moins une sous-séquence répétée dans ladite séquence, et
- 5 c) on évalue des instants de début et de fin de ladite sous-séquence dans la séquence sonore.

Avantageusement, selon une étape supplémentaire :

- 10 d) on extrait la sous-séquence précitée pour stocker, dans une mémoire, des échantillons sonores représentant ladite sous-séquence.

Préférentiellement, l'extraction de l'étape d) concerne au moins une sous-séquence dont la durée est la plus importante et/ou une sous-séquence dont la fréquence de répétition est la plus importante dans ladite séquence.

La présente invention trouve une application avantageuse dans l'aide à la détection de défaillances de machines industrielles ou de moteurs, notamment en obtenant des séquences d'enregistrement sonore de phases d'accélération et de décélération du régime moteur. L'application du procédé au sens de l'invention permet d'isoler une sous-séquence sonore correspondant par exemple à un plein régime ou à une phase d'accélération, cette sous-séquence étant, le cas échéant, comparée à une sous-séquence de référence.

30 Dans une autre application avantageuse à l'obtention de données musicales du type décrit ci-avant, la séquence

sonore précitée est un morceau de musique comportant une succession de sous-séquences parmi au moins une introduction, un couplet, un refrain, un pont de transition, un thème, un motif, ou un mouvement qui se répète dans la séquence.

A l'étape c), on détermine préférentiellement au moins les instants respectifs de début et de fin d'une première sous-séquence et d'une seconde sous-séquence.

Dans une réalisation particulièrement avantageuse, à l'étape d), on extrait alors une première et une seconde sous-séquence pour obtenir, sur un support mémoire, un résumé sonore dudit morceau de musique comportant au moins la première sous-séquence enchaînée avec la seconde sous-séquence.

Préférentiellement, la première sous-séquence correspond à un couplet et la seconde sous-séquence correspond à un refrain.

Toutefois, il peut advenir que des première et seconde sous-séquences, extraites d'une séquence sonore, ne soient pas contiguës dans le temps.

A cet effet, on prévoit en outre les étapes suivantes :

d1) détecter au moins une cadence de la première sous-séquence et/ou de la seconde sous-séquence pour estimer la durée moyenne d'une mesure à ladite cadence, ainsi qu'au moins un segment de fin de la première sous-séquence et au moins un segment de début de la seconde sous-séquence, de durées respectives correspondant sensiblement à ladite

durée moyenne et isolés dans la séquence d'un nombre entier de durées moyennes,  
d2) générer au moins une mesure de transition de durée correspondant à ladite durée moyenne et comportant une addition des échantillons sonores d'au moins ledit segment de fin et d'au moins ledit segment de début,  
d3) et concaténer la première sous-séquence, la ou les mesures de transition et la seconde sous-séquence pour obtenir l'enchaînement de la première et de la seconde sous-séquence.

On constatera que la succession des étapes d1) à d3) trouve, au-delà de la génération automatique de résumés sonores, une application avantageuse à la création musicale assistée par ordinateur. Dans cette application, un utilisateur peut créer lui-même deux sous-séquences d'un morceau musical, tandis qu'un logiciel comportant des instructions pour dérouler les étapes d1) à d3) assure un enchaînement des deux sous-séquences par concaténation, sans artéfact et agréable à l'oreille.

De façon plus générale, la présente invention vise aussi un produit programme d'ordinateur, stocké dans une mémoire d'ordinateur ou sur un support amovible propre à coopérer avec un lecteur homologue d'ordinateur, et comportant des instructions pour dérouler les étapes du procédé au sens de l'invention.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après, et des dessins annexés sur lesquels :

- la figure 1a représente un signal audio d'un morceau de musique correspondant, dans l'exemple représenté, à une chanson de variété ;
- la figure 1b représente la variation d'énergie spectrale en fonction du temps, pour le morceau de musique dont le signal audio est représenté sur la figure 1a ;
- la figure 1c illustre les durées occupées par les différents passages du morceau de musique de la figure 1a et qui se répètent dans ce morceau ;
- la figure 2 représente schématiquement des fenêtres temporelles sélectionnées dans deux parties respectives du morceau musical pour préparer la concaténation de ces deux parties, selon la succession des étapes d1) à d3) ci-avant ;
- la figure 3a représente schématiquement des segments  $s_i(t)$  et  $s_j(t)$  sélectionnés dans les parties respectives du morceau précitées, pour préparer une concaténation des deux parties par superposition/addition ;
- la figure 3b illustre schématiquement par le signe " $\oplus$ " la superposition/addition précitée ;
- la figure 4 illustre une fenêtre temporelle pour la concaténation précitée, de forme et de largeur préférées ; et
- la figure 5 représente un organigramme de traitement d'une séquence sonore, dans un mode de réalisation préféré de la présente invention.

Le signal audio de la figure 1a représente l'intensité sonore (en ordonnées) en fonction du temps (en abscisses)

d'un morceau musical (ici, le morceau "head over feet"<sup>®</sup> de l'artiste Alanis Morissette). Pour construire ce signal audio, les signaux respectifs des voies droite et gauche (en mode stéréophonique) ont été synchronisés et additionnés.

Au signal audio représenté sur la figure 1a, on applique une transformée spectrale (par exemple de type transformée de Fourier rapide FFT) pour obtenir une variation temporelle de l'énergie spectrale du type représenté sur la figure 1b.

Dans un mode de réalisation, il s'agit d'une pluralité de FFT à court terme, successives, dont le résultat est appliqué à un banc de filtres sur plusieurs gammes de fréquences (préférentiellement de largeurs de bande croissantes comme le logarithmique de la fréquence). Une autre transformée de Fourier est ensuite appliquée pour obtenir des paramètres dynamiques du signal audio (référéncés PD sur la figure 1b). En particulier, l'échelle en ordonnées de la figure 1b indique l'amplitude des variations des composantes à différentes vitesses dans un domaine de fréquences donné. Ainsi, l'indice 0 ou 2 de l'échelle arbitraire en ordonnées de la figure 1b correspond à une variation lente dans les basses fréquences, tandis que l'indice 12 de cette même échelle correspond à une variation rapide dans les hautes fréquences. Ces variations sont exprimées en fonction du temps, en abscisses (secondes). Les intensités associées à ces paramètres dynamiques PD, au cours du temps, sont illustrées par différents niveaux de gris dont les valeurs



relatives sont indiquées par la colonne de référence COL (à droite de la figure 1b).

On indique que les paramètres dynamiques du type représenté sur la figure 1b permettent d'identifier complètement un morceau de musique. Dans ce contexte d'"empreinte" d'un morceau de musique, la demande FR-2834363 de la Demanderesse décrit de façon détaillée ces paramètres et la manière de les obtenir.

En variante, les variables déduites du signal audio et permettant de caractériser le morceau de musique peuvent être de type différent, notamment des coefficients dits "Mel Frequency Cepstral Coefficients". Globalement, on indique que ces coefficients (connus en soi) sont encore obtenus par transformée de Fourier rapide, à court terme.

La figure 1c offre une représentation visuelle de l'évolution de l'énergie spectrale de la figure 1b. Sur la figure 1c, l'abscisse représente le temps (en secondes) et les ordonnées représentent les différentes parties du morceau, tels que les couplets, les refrains, l'introduction, un thème, ou autres. La répétition au cours du temps d'une partie similaire, tel qu'un couplet ou un refrain, est représentée par des rectangles grisés qui apparaissent à différentes abscisses dans le temps (et qui peuvent être de largeurs temporelles différentes), mais de même ordonnées. Pour passer de la représentation de la figure 1b à la représentation de la figure 1c, on met en œuvre une analyse statistique en utilisant par exemple l'algorithme des "K-means", ou encore l'algorithme

"FUZZY K-means", ou encore une chaîne de Markov cachée, avec un apprentissage par l'algorithme BAUM-WELSH, suivi d'une évaluation par l'algorithme de VITERBI.

5 Typiquement, la détermination du nombre d'états (les parties du morceau de musique) qui sont nécessaires à la représentation d'un morceau de musique s'effectue de façon automatisée, par comparaison de la similarité des états trouvés à chaque itération des algorithmes précités, et en  
10 éliminant les états redondants. Cette technique, dite de "pruning" permet ainsi d'isoler chaque partie redondante du morceau de musique et de déterminer ses coordonnées temporelles (ses instants de début et de fin, comme indiqué ci-avant).

15

Ainsi, on étudie les variations, par exemple dans les fréquences tonales (d'une voix humaine), de l'énergie spectrale pour déterminer la répétition d'un passage musical particulier dans le signal audio.

20

Préférentiellement, on cherche à extraire un ou plusieurs passages musicaux dont la durée est la plus importante dans le morceau de musique et/ou dont la fréquence de répétition est la plus importante.

25

Par exemple, pour la plupart des morceaux de variété, on pourra choisir d'isoler les parties de refrain, dont la répétition est généralement la plus fréquente, puis les parties de couplet, dont la répétition est fréquente,  
30 puis, le cas échéant, d'autres parties encore si elles se répètent.

On indique que d'autres types de sous-séquences représentatives du morceau de musique peuvent être extraites, dès lors que ces sous-séquences se répètent dans le morceau de musique. Par exemple, on peut choisir d'extraire un motif musical, généralement de plus courte durée qu'un couplet ou un refrain, tel qu'un passage de percussions répété dans le morceau de musique, ou encore une phrase vocale scandée plusieurs fois dans le morceau. En outre, un thème peut aussi être extrait du morceau de musique, par exemple une phrase musicale répétée dans un morceau de jazz ou de musique classique. En musique classique, un passage tel qu'un mouvement peut en outre être extrait.

Sur le résumé visuel représenté à titre d'exemple sur la figure 1c, les rectangles grisés indiquent la présence d'une partie du morceau telle que l'introduction ("intro"), d'un couplet ou d'un refrain dans une fenêtre temporelle indiquée par l'abscisse temporelle (en secondes). Ainsi, entre 0 et environ 15 secondes, le morceau de musique démarre par une introduction (indexée par le chiffre 2 sur l'échelle en ordonnées). L'introduction est suivie de deux alternances de couplet (indexé par le chiffre 3) et de refrain (indexé par le chiffre 1) jusqu'à environ 100 secondes.

On se réfère maintenant à la figure 5 pour décrire les étapes principales du procédé pour obtenir le résumé sonore précité, selon un mode de réalisation préféré. Tout d'abord, on obtient les signaux audio sur la voie gauche

"audio L" et sur la voie droite "audio R" aux étapes respectives 10 et 11, lorsque la séquence sonore initiale est représentée en mode stéréophonique. Les signaux de ces deux voies sont additionnés à l'étape 12 pour obtenir un  
5 signal audio du type représenté sur la figure 1a. Ce signal audio est, le cas échéant, stocké sous forme échantillonnée dans une mémoire de travail avec des valeurs d'intensité sonores rangées en fonction de leurs coordonnées temporelles associées (étape 14). A ces  
10 données audio, on applique une transformée spectrale (de type FFT dans l'exemple représenté), à l'étape 16, pour obtenir, à l'étape 18, les coefficients spectraux  $F_i(t)$  et/ou leur variation  $\Delta F_i(t)$  en fonction du temps. A l'étape 20, un module d'analyse statistique opère sur la  
15 base des coefficients obtenus à l'étape 18 pour isoler des instants  $t_0, t_1, \dots, t_7$  qui correspondent à des instants de début et de fin des différentes sous-séquences qui se répètent dans le signal audio de l'étape 14.

20 Dans l'exemple représenté, le morceau de musique présente une structure (classique en variété) du type comportant :

- une introduction dans le début du morceau entre un instant  $t_0$  et un instant  $t_1$ ,
- un couplet entre  $t_1$  et  $t_2$ ,
- 25 - un refrain entre  $t_2$  et  $t_3$ ,
- un second couplet entre  $t_3$  et  $t_4$ ,
- un second refrain entre  $t_4$  et  $t_5$ ,
- une introduction, à nouveau, le cas échéant agrémentée d'un solo instrumental, entre les instants  
30  $t_5$  et  $t_6$ , et

- la répétition de deux refrains de fin de morceau entre les instants  $t_6$  et  $t_7$ .

5 A l'étape 22, les instants  $t_0$  à  $t_7$  sont répertoriés et indexés en fonction du passage musical correspondant (introduction, couplet ou refrain) et stockés, le cas échéant, dans une mémoire de travail. A l'étape 23, on peut construire alors un résumé visuel de ce morceau de musique, tel que représenté sur la figure 5.

10 Dans l'exemple décrit ci-avant d'un morceau de variété comportant une structure typique, on construit le résumé sonore à partir d'un couplet extrait du morceau, suivi d'un refrain extrait du morceau. A l'étape 24, on prépare  
15 une concaténation des échantillons sonores du signal audio entre les instants  $t_1$  et  $t_2$ , d'une part, et entre les instants  $t_2$  et  $t_3$ , d'autre part, dans l'exemple décrit. Le cas échéant, le résultat de cette concaténation est mémorisé dans une mémoire permanente MEM pour une  
20 utilisation ultérieure, à l'étape 26.

Toutefois, en règle générale, l'instant de fin d'un couplet isolé et l'instant de début d'un refrain isolé ne sont pas forcément identiques, ou encore, on peut choisir  
25 de construire le résumé sonore à partir du premier couplet et du second refrain (entre  $t_4$  et  $t_5$ ) ou du refrain de fin (entre  $t_6$  et  $t_7$ ). Ainsi, les deux passages sélectionnés pour construire le résumé sonore ne sont pas forcément contigus.

Une concaténation aveugle de signaux sonores correspondant à deux parties d'un morceau de musique donne un ressenti désagréable à l'oreille. On décrit ci-après, en référence aux figures 2, 3a, 3b et 4, la construction d'un signal sonore par concaténation de deux parties d'un morceau de musique, de manière à surmonter ce problème.

L'un des buts de cette construction par concaténation est de préserver localement le tempo du signal sonore.

Un autre but est d'assurer une distance temporelle entre points de concaténation (ou points d'"alignement") égale à un multiple entier de la durée d'une mesure.

Préférentiellement, cette concaténation s'effectue par superposition/addition de segments sonores choisis et isolés des deux parties respectives précitées du morceau de musique.

On décrit ci-après une superposition/addition de tels segments sonores, dans un premier temps par synchronisation de beat (dite "beat-synchrone"), puis par synchronisation de mesure selon une réalisation préférée.

On note ci-après :

- *bpm*, le nombre de beats par minute d'un morceau de musique,
- *D*, la référence de ce nombre *bpm* (par exemple dans le cas d'un morceau noté "120=noire", *bpm*=120 et *D*=noire),

- T, la durée (exprimée en secondes) d'un beat, c'est-à-dire de la référence D : dans l'exemple précédent où D=noire, on a

$$T = \frac{60}{bpm}$$

- 5 - N, le numérateur de la métrique du morceau de musique (par exemple, dans le cas d'une mesure notée "3/4", N=3),
- M, la durée (exprimée en secondes) d'une mesure, donnée par la relation  $M=N.T$  (soit  $M=3*60/120$  dans l'exemple précédent),
- 10 -  $s(t)$ , le signal audio d'un morceau de musique,
- $\hat{s}(t)$ , le signal reconstruit par superposition/addition, et
- $s_i(t)$  et  $s_j(t)$ , les  $i^{\text{ème}}$  et  $j^{\text{ème}}$  segments qui
- 15 comportent des signaux audio respectifs appartenant à un premier et à un second passage d'un morceau de musique, et qui sont utilisés pour la construction de  $\hat{s}(t)$  par superposition/addition.

- 20 En principe, les premier et second passages précités ne sont pas contigus.  $\hat{s}(t)$  est alors obtenu comme suit.

En se référant à la figure 2, les segments  $s_i(t)$  et  $s_j(t)$  sont d'abord formés par découpage du signal audio à l'aide

25 d'une fenêtre temporelle  $h_L(t)$ , de largeur L et définie (de valeur non nulle) entre 0 et L. Cette fenêtre peut être de type rectangulaire, de type dit "de hanning", de type dit "de hanning en palier", ou autre. En se référant à la figure 4, un type préféré de fenêtre temporelle est

30 obtenu par concaténation d'un flanc montant, d'un pallier

et d'un flanc descendant. La largeur temporelle préférée de cette fenêtre est indiquée ci-après.

On définit alors le premier segment  $s_i(t)$  de sorte que :

$$s_i(t) = s(t + m_i) \cdot h_L(t) \quad [1]$$

où  $m_i$  est l'instant de début du premier segment.

Comme le montre la figure 3a, on construit  $s_j(t)$  sensiblement de la même manière :

$$s_j(t) = s(t + m_j) \cdot h_L(t) \quad [1bis]$$

où  $m_j$  est l'instant de début du second segment.

Même si la durée  $L$  de la fenêtre temporelle est la même pour les deux segments, on indique toutefois que la forme de la fenêtre peut être différente d'un segment  $s_i(t)$  à l'autre  $s_j(t)$ , comme le montre d'ailleurs la figure 2.

Soit  $b_i$  et  $b_j$  deux positions respectives à l'intérieur des premier et second segments, appelées "positions de synchronisation", par rapport auxquelles s'effectue la superposition/addition, telles que :

$$0 \leq b_i \leq L \quad \text{et} \quad 0 \leq b_j \leq L \quad [2]$$

Avantageusement, la distance temporelle entre  $b_i$  et  $b_j$  est choisie égale à un multiple entier de la durée  $T$  d'un beat ( $b_j - b_i = kT$ ). Dans ces conditions, on dit qu'il y a reconstruction "beat-synchrone" si

$$\hat{s}(t) = \sum_i s'_i(t - (i-1) \cdot (k'T) + c) \quad [4]$$

$$\text{avec} \quad s'_i(t) = s_i(t + b_i) \quad [5]$$



et où  $k'$  est le plus grand entier tel que  $k'T \leq L - (b_i - m_i)$ ,  
 $c$  est une constante de temps telle que  $c = b_i - m_i$ .  
 Avantageusement, la distance entre les instants  $m_i$  et  $m_j$   
 est choisie égale à un multiple entier de  $k'NT$ , dans  
 5 lequel  $N$  désigne le numérateur de la métrique.

Ainsi, le signal reconstruit s'écrit :

$$\hat{s}(t) = \sum_i s_i'(t - (i-1) \cdot (k'NT) + c)$$

10 On obtient alors une superposition/addition synchrone à la  
 mesure. La figure 3b illustre cette situation. On remarque  
 sur la figure 4 que la largeur  $L$  de la fenêtre temporelle  
 précitée est voisine de  $k'NT$  (aux flancs montant et  
 descendant près). Toutefois, on choisira  
 15 préférentiellement dans ce cas des rampes de flancs telles  
 que  $k'T \leq L - 2(b_i - m_i)$ .

Plus particulièrement, on choisit les instants  $m_i$  et  $m_j$  de  
 sorte qu'ils correspondent à des premiers temps de  
 20 mesures. Dans ces conditions, on obtient avantageusement  
 une superposition/addition beat-synchrone dite "alignée".

Ainsi, en déterminant en outre la métrique du premier  
 passage et/ou du second passage, on peut effectuer une  
 25 reconstruction beat-synchrone à la mesure. Si, de plus, on  
 choisit les premier et second segments de sorte qu'ils  
 commencent par un premier temps de mesure, cette  
 reconstruction beat-synchrone est alignée.

On indique qu'une reconstruction du signal  $\hat{s}(t)$  peut être menée sur la base de plus de deux passages musicaux à concaténer. Pour  $i$  passages musicaux ( $i > 2$ ), la généralisation du procédé ci-avant s'exprime par la relation :

$$\hat{s}(t) = s_1'(t+c) + s_2'(t-k_1'T+c) + s_3'(t-k_1'T+k_2'T+c) + \dots \\ + s_i'(t + \sum_{j=1}^i (-1)^j k_j'T + c)$$

Chaque entier  $k_j'$  est défini comme l'entier le plus grand tel que  $k_j'T \leq L_j - (b_j - m_j)$ , où  $L_j$  correspond à la largeur de la fenêtre du jème passage musical à concaténer.

On indique que les premiers temps de mesure, ou encore la métrique, ou encore le tempo d'un morceau de musique, peuvent être détectés automatiquement, par exemple en utilisant des applications logicielles existantes. Par exemple, la norme MPEG-7 (Audio Version 2) prévoit la détermination et la description du tempo et de la métrique d'un morceau de musique, en utilisant de telles applications logicielles.

Bien entendu, la présente invention ne se limite pas à la forme de réalisation décrite ci-avant à titre d'exemple ; elle s'étend à d'autres variantes.

Ainsi, on comprendra que le résumé sonore peut comporter plus de deux passages musicaux, par exemple une introduction, un couplet et un refrain, ou encore deux

passages différents d'un couplet et d'un refrain, tels que l'introduction et un refrain, par exemple.

5 On remarquera aussi que les étapes représentées sous forme d'organigramme sur la figure 5 peuvent être implémentées par un logiciel informatique dont l'algorithme reprend globalement la structure de l'organigramme. A ce titre, la présente invention vise aussi un tel programme informatique.

Revendications

1. Procédé de traitement d'une séquence sonore, dans lequel :

5 a) on applique une transformée spectrale à ladite séquence pour obtenir des coefficients spectraux variant en fonction du temps dans ladite séquence,

caractérisé en ce qu'il comporte en outre les étapes suivantes :

10 b) on détermine, par analyse statistique desdits coefficients spectraux, au moins une sous-séquence répétée dans ladite séquence, et

c) on évalue des instants de début et de fin de ladite sous-séquence dans la séquence sonore.

15

2. Procédé selon la revendication 1, caractérisé en ce qu'il comporte en outre une étape :

20 d) d'extraction de la sous-séquence pour stocker, dans une mémoire, des échantillons sonores représentant ladite sous-séquence.

3. Procédé selon la revendication 2, caractérisé en ce que l'extraction de l'étape d) concerne au moins une sous-séquence dont la durée est la plus importante et/ou une  
25 sous-séquence dont la fréquence de répétition est la plus importante dans ladite séquence.

4. Procédé selon l'une des revendications 1 à 3, dans lequel la séquence sonore est un morceau de musique  
30 comportant une succession de sous-séquences parmi au moins

une introduction, un couplet, un refrain, un pont de transition, un thème, un motif, un mouvement,

caractérisé en ce que, à l'étape c), on détermine au moins les instants respectifs de début et de fin d'une première sous-séquence et d'une seconde sous-séquence.

5. Procédé selon la revendication 4, prise en combinaison avec la revendication 3, caractérisé en ce que la première sous-séquence correspond à un couplet et la seconde sous-séquence correspond à un refrain.

6. Procédé selon l'une des revendications 4 et 5, prises en combinaison avec la revendication 2, caractérisé en ce que, à l'étape d), on extrait une première et une seconde sous-séquence pour obtenir, sur un support mémoire, un résumé sonore dudit morceau de musique comportant au moins la première sous-séquence enchaînée avec la seconde sous-séquence.

7. Procédé selon la revendication 6, dans lequel les extraits des sous-séquences sont non contigus dans le temps, caractérisé en ce qu'il comporte en outre les étapes suivantes :

d1) détecter au moins une cadence de la première sous-séquence et/ou de la seconde sous-séquence pour estimer la durée moyenne d'une mesure à ladite cadence, ainsi qu'au moins un segment de fin de la première sous-séquence et au moins un segment de début de la seconde sous-séquence, de durées respectives correspondant sensiblement à ladite durée moyenne et isolés dans la séquence d'un nombre entier de durées moyennes,

d2) générer au moins une mesure de transition de durée correspondant à ladite durée moyenne et comportant une addition des échantillons sonores d'au moins ledit segment de fin et d'au moins ledit segment de début,

5 d3) et concaténer la première sous-séquence, la ou les mesures de transition et la seconde sous-séquence pour obtenir un enchaînement de la première et de la seconde sous-séquence.

10 8. Procédé selon la revendication 7, caractérisé en ce que l'étape d1) comporte un découpage en au moins deux fenêtres, de type rectangulaire, de type de Hanning, de type de Hanning en palier, ou préférentiellement de type comportant un flanc montant, un pallier et un flanc  
15 descendant dans le temps.

9. Procédé selon l'une des revendications 7 et 8, caractérisé en ce que l'étape d2) comporte une reconstruction beat-synchrone.

20

10. Procédé selon la revendication 9, caractérisé en ce que, à l'étape d1), on détermine en outre la métrique de la première sous-séquence et/ou de la seconde sous-séquence, et en ce que l'étape d2) comporte une  
25 reconstruction beat-synchrone à la mesure.

11. Procédé selon l'une des revendications 9 et 10, caractérisé en ce que, à l'étape d1), on détermine lesdits segments de fin et de début de sorte qu'ils commencent par  
30 un premier temps de mesure, et en ce que l'étape d2) comporte une reconstruction beat-synchrone alignée.

12. Produit programme d'ordinateur, stocké dans une  
mémoire d'ordinateur ou sur un support amovible propre à  
coopérer avec un lecteur d'ordinateur, caractérisé en ce  
5 qu'il comporte des instructions pour dérouler les étapes  
du procédé selon l'une des revendications précédentes.

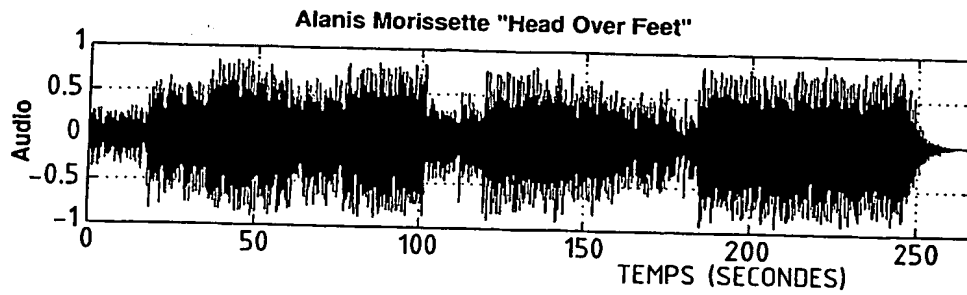


FIG. 1a

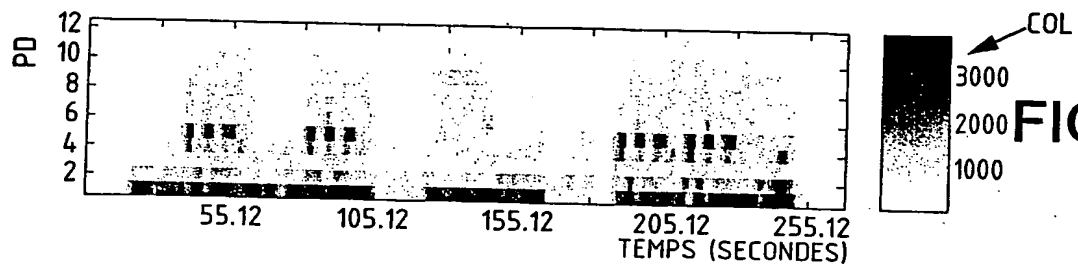


FIG. 1b

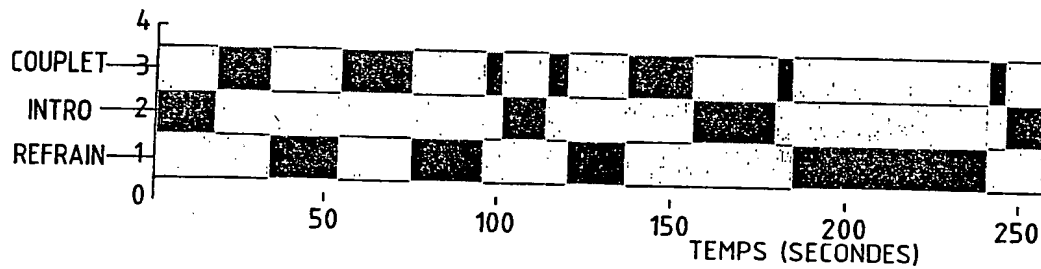


FIG. 1c

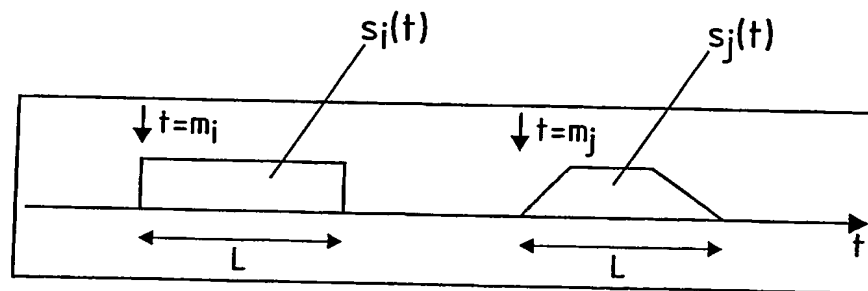


FIG. 2



2/3

